UNIVERSITI PUTRA MALAYSIA

GENERATION OF EXPRESSED SEQUENCE TAGS (ESTs) FROM NON-EMBRYOGENIC CALLUS cDNA LIBRARIES OF OIL PALM (*ELAEIS GUINEENSIS* JACQ.)

AMOSTAN CHI YEE

FSMB 2003 10

GENERATION OF EXPRESSED SEQUENCE TAGS (ESTs) FROM NON-EMBRYOGENIC CALLUS cDNA LIBRARIES OF OIL PALM *(ELAEIS GUINEENSIS* JACQ.*)*

By

AMOS TAN CHI YEE

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of Master of Science

July 2003

Dedicated to Dad, Mom, Audrey,
Adrian and Yoke Fah

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirements for the degree of Master of Science

GENERATION OF EXPRESSED SEQUENCE TAGS (ESTs) FROM NON-EMBRYOGENIC CALLUS cDNA LIBRARIES OF OIL PALM *(ELAEIS GUINEENSIS* JACQ.*)*

By

AMOS TAN CHI YEE

July 2003

Chairman : Dr. Tan Siang Hee

Faculty : Food Science and Biotechnology

Due to the economic importance of tissue culture in oil palm, it is desirable to understand the chemical characteristics and molecular changes associated with callogenesis and embryogenesis. Therefore partial sequencing of randomly selected complementary DNA (cDNA) clones was carried out to investigate and understand the gene expression pattern during non-embryogenic callus in oil palm. A total of 2,296 clones from cDNA libraries constructed using messenger RNA (mRNA) from non-embryogenic callus tissues have been partially sequenced to generate expressed sequence tags (ESTs). The comparison of the coding capacity of these 1,832 ESTs with the non-redundant sequences in the National Center for Biotechnology Information (NCBI) databases resulted in assignment of putative function to 971 (53.0%) of the ESTs. The remaining 861 (47.0%) were considered as novel genes. Hybridization of the 80 selected cDNA clones to ^{32}P-labelled first strand cDNA probes synthesized from the total mRNA of NEC was compared with probes prepared from mRNA of leaves. A total of 12 clones, which hybridized predominantly to the NEC but not to the leaf probe were selected for Northern analysis. The Northern analysis showed that 9 out of the 12 ESTs clones

selected were expressed predominantly in the callus tissue. However, carrying out a sequence comparison using the BLAST algorithm, only 1 out of the 9 clones had significant homology to a known gene, glutathione-S-transferase (GST) gene. The remaining 8 unidentifiable ESTs were then subjected to a motif search using the Blocks IMProved Searcher (BLIMPS) program. Based on the search results, 4 different motifs were identified to be related to the 8 ESTs and they are either involved in developmental regulation or protein storage function. The results presented here therefore demonstrate the effectiveness and efficiency of the EST strategy in gene discovery in the oil palm callus tissue.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Master Sains

PENJANAAN PENANDA JUJUKAN TEREKSPRES (ESTs) DARIPADA KALUS TAK-EMBRIOGENIK PERPUSTAKAAN cDNA POKOK SAWIT *(ELAEIS GUINEENSIS* JACQ.*)*

Oleh

AMOS TAN CHI YEE

Julai 2003

Pengerusi : Dr. Tan Siang Hee

Fakulti : Sains Makanan dan Bioteknologi

Berdasarkan kepada kepentingan ekonomi kultur tisu pokok sawit, adalah menjadi satu keperluan untuk memahami pertukaran ciri-ciri kimia dan molekul yang berkaitan dengan proses kalogenesis dan embriogenesis. Justeru itu, penjujukan-separa klon-klon DNA pelengkap (cDNA) yang dipilih secara rawak telah dijalankan bagi memahami paten pengekspresan gen-gen pada kalus tak embriogenik (NEC) pokok sawit. Perpustakaan cDNA telah dibina menggunakan RNA pengutus (mRNA) daripada kalus tak embriogenik. Sebanyak 2,296 klon daripada perpustakaan cDNA telah dijujukkan untuk mengenalpasti penanda jujukan terekspres (EST). Perbandingan kapasiti pengkodan 1,832 EST dengan jujukan tak redundan di pangkalan data National Center for Biotechnology Information (NCBI) telah mengenalpasti 917 (53.0%) EST mempunyai fungsi putatif. Baki 861 (47.0%) EST dianggap sebagai gen-gen novel. Sebanyak 80 cDNA klon yang telah terpilih dihibridisasi kepada prob cDNA dimana bebenang pertamanya dilabel dengan ^{32}P. Prob cDNA disintesis daripada mRNA NEC dan dibandingkan dengan prob-prob yang dihasilkan daripada mRNA daun. Sejumlah 12 klon telah menghibridisasi secara dominan kepada NEC tetapi tidak kepada prob-prob daun yang telah dipilih untuk analisis Northern. Analisis Northern menunjukkan 9

daripada 12 klon EST mengekspres secara dominan dalam tisu kalus. Bagaimanapun, perbandingan jujukan menggunakan algoritma BLAST menunjukkan hanya 1 daripada 9 klon menpunyai homologi yang signifikan kepada gen yang dikenali iaitu gen glutathione-S-transferase. Lapan EST yang tidak dapat dikenalpasti telah dianalisis menerusi pencarian motif menggunakan program BLIMPS (Blocks IMProved Searcher). Berdasarkan kepada keputusan yang diperolehi, 4 motif yang berlainan telah dikenalpasti mempunyai kaitan dengan 8 EST tersebut. Motif-motif tersebut terlibat sama ada dalam fungsi regulasi pengembangan atau penyimpanan protein. Keputusan-keputusan yang diperolehi menunjukkan keberkesanan strategi EST bagi penemuan gen-gen dalam kalus pokok sawit.

ACKNOWLEDGEMENTS

I wish to express my sincere appreciation and heartfelt gratitude to Dr. Tan Siang Hee, Dr. Cheah Suan Choo, Dr. Sharifah Shahrul and Dr. Suhaimi Napis for their invaluable advice and guidance throughout the course of this project as my supervisors. I'm indebted to Mr. Rajinder for his ideas, support and unconditional help that surpass his responsibility as a group leader. Furthermore I'm blessed to have him as a friend who shares the same 'highly evolved' and 'God-forbid' sense humour as mine. I'm also forever grateful to Dr. Meilina Ong (fondly known as Mei) for literally handing over to me 'her bible of knowledge' that was actually the nucleus of this 'Pulp Fiction' of mine. My appreciation also extends to Dr. Ravigadevi, Dr. Arif, Dr. Maizura, Dr. Siti, Dr. Maria and Dr. Tarmizi for contributing in many different ways. To Mdm Wooi (formerly at Ebor Laboratories) and Pn. Suzaini (Ebor Laboratories), the deepest appreciation for their constant supply of oil palm callus and embryoid tissue, without which, this project would not be progressing as smoothly as it did. I also wish to thank the Ministry of Science, Technology and the Environment (MOSTE) for funding this research, under the Malaysia-MIT Biotechnology Partnership Programme (MMBPP).

An extra thanks also goes out to Dr. Cheah for employing me in the first place under the MMBPP programme. It was there that I spent a memorable 2 ½ years of my life with a bunch of outstanding 'rebels', who I now consider my "Band of Brothers" (Kia Ling, Ah Ti, Andy, CK, Tuck Fatt, Jihan) and "Sisters" (Halimah, Elyana, Soo Heong, Pek Lan, Mel, Shah, Jane, Nina, Emily & Noni). Many of you shared your life's colourful portraits, technical expertise, and resources that I'll be indebted to forever. I'm also fortunate to have the initial 'jump-start' provided by the lovely tissue culturist upstairs (Mei, Dr. Sharifah, Siew Eng, Parames & Ayu) with the classification of my

EST clones. I'm also blessed to have friends like Weng Wah (bioinformatics guru), Kia Ling and Chieh Wen (blotting wizards) who provided me FOC with their highly sought after skills and advice, without which my project would have taken virtually forever to complete.

The unsung heroes, the ARO's (Ms. Leslie Ooi, Ms. Rahimah, En. Jamil,, & Faizun) the RA's under the genomics group (particularly, Siti, Abang Cat, Abang Zul, Mat Nor, Mat Shari, Kak. Jabariah, Ayu, Shikin, and etc.) and the administrative staff (Kak. Wai, Rahaya and etc) who helped me in more ways than one that would be almost impossible to catalog them all. Thank you all for also sharing some colourful moments of your life with me while I was constantly waiting for the PCR, centrifuge or the vacuum concentrator to finish processing my samples. Special thanks also to my housemates (Chen, Leslie, Andy) for safe-guarding my earthly processions and biological organisms (my feng-shui goldfishes) to the best of their abilities while I was living out of my Samsonite cabin luggage (55x40x20cm) for the last 11 months. Thank you guys also for making sure that I'm always well fed and handling the administrative matters associated with the 'twilight zone policies' of our graduate school. What success that I have today I owe a great deal to my family and my friend, Yee Joo who had to sacrifice her many weekends to accompany me to the laboratory for which the boredom would almost be unbearable without a good-spirited person like her. Last but not the least, I would like to thank my better half, Yoke Fah for being there when I needed it most and sacrificing so much for so little. Thank you all for your support and unconditional love.

I certify that an Examination Committee met on 18th June 2003 to conduct the final examination of Amos Tan Chi Yee on his Master of Science thesis entitled "Generation of Expressed Sequence Tags (ESTs) from Non-embryogenic Callus cDNA Libraries of Oil Palm (*Elaeis guineensis* Jacq.)" in accordance with Universiti Pertanian Malaysia (Higher Degree) Act 1980 and Universiti Pertanian Malaysia (Higher Degree) Regulations 1981. The Committee recommends that the candidate be awarded the relevant degree. Members of the Examination Committee are as follows:

MICHAEL WONG VUI LING, Ph.D.,
Faculty of Food Science and Biotechnology
Universiti Putra Malaysia
(Chairman)

TAN SIANG HEE, Ph.D.,
Lecturer
Faculty of Food Science and Biotechnology
Universiti Putra Malaysia
(Member)

SUHAIMI NAPIS, Ph.D.,
Assoc. Professor/ Deputy Director
Institute Multimedia and Software
Universiti Putra Malaysia
(Member)

CHEAH SUAN CHOO, Ph.D.,
Head
Advanced Breeding & Biotechnology Centre
Biology Division
Malaysia Palm Oil Board
(Member)

SHARIFAH SHARUL RABIAH SYED ALWI, Ph.D.,
Senior Research Officer
Advanced Breeding & Biotechnology Centre
Biology Division
Malaysia Palm Oil Board
(Member)

GULAM RUSUL RAHMAT ALI, Ph.D.
Professor/Deputy Dean
School of Graduate Studies,
Universiti Putra Malaysia

Date: 4 SEP 2003

This thesis submitted to the Senate of Universiti Putra Malaysia has been accepted as partial fulfillment of the requirements for the degree of Master of Science. The members of the Supervisory Committee are as follows.

TAN SIANG HEE, Ph.D.,
Lecturer
Faculty of Food Science and Biotechnology
Universiti Putra Malaysia
(Chairman)

SUHAIMI NAPIS, Ph.D.,
Assoc. Professor/ Deputy Director
Institute Multimedia and Software
Universiti Putra Malaysia
(Member)

CHEAH SUAN CHOO, Ph.D.,
Head
Advanced Breeding & Biotechnology Centre
Biology Division
Malaysia Palm Oil Board
(Member)

SHARIFAH SHARUL RABIAH SYED ALWI, Ph.D.,
Senior Research Officer
Advanced Breeding & Biotechnology Centre
Biology Division
Malaysia Palm Oil Board
(Member)

AINI IDERIS, Ph.D.,
Professor/Dean,
School of Graduate Studies,
Universiti Putra Malaysia

Date: 16 SEP 2003

DECLARATION

I hereby declare that the thesis is based on my original work except for quotations, and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at UPM or other institutions

Jmosbar .

AMOS TAN CHI YEE

Date: 6/10/02

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

%	percentage
α	alpha
β	beta
λ	lambda
°C	degree Celsius
μg	microgram
μl	microliter
μM	micromolar
A	Adenine
BIS	N'N'-methylenebisacrylamide
BLAST	Basic Local Alignment Search Tool
bp	base pair
C	Cytosine
$CaCl_2$	Calcium Chloride
cDNA	copy DNA
cfu	colony forming unit
Ci	Curie
cm	centimeter
cps	counts per second
CsCl	Cesium Chloride
C-terminal	Carboxyl terminal
D x P	*Dura* x *Pisifera*
dATP	2'-deoxy-adenosine-5'-triphosphate
dCTP	2'-deoxy-cytidine-5'-triphosphate

DEPC	Diethyl Pyrocarbonate
dGTP	2'-deoxy-guanosine-5'-triphosphate
dH_2O	deionized water
DMF	Dimethyl Fluoride
DMSO	Dimethylsulphonyl oxide
DNA	Deoxyribonucleic acid
Dnase 1	Deoxyribonuclease 1
dNTP	deoxynucleotide triphosphates
dTTP	2'-deoxy-thymidine-5'-triphosphate
E.coli	*Esherichia coli*
EC	Embryogenic Callus
e.g.	example
EDTA	Ethylenediaminetetraacetatic acid
EGTA	Ethylene glycol bis- (β-aminoethylene ether)
Em	Erythromycin
ESTs	Expressed Sequences Tags
EtBr	Ethidium Bromide
g	gram
G	Guanine
HCl	Hydrochloride Acid
hr	hours
IPTG	isopropyl-β-D-thiogalactoside
Jacq.	*Jacquin*
k	kilo
kb	kilobase

KCl	Potassium Chloride
kDA	Kilodalton
L	Liter
LB	Luria Bertani
LiCl	Lithium Chloride
M	Molar
mA	milliampere
$MgCl_2$	Magnesium Chloride
$MgSO_4$	Magnesium Sulphate
min	minute
ml	milliliter
mm	millimeter
mM	millimolar
MMLV	Maurine Moloney Leukemia Virus
mmol	millimole
MPOB	Malaysia Palm Oil Board
mRNA	messenger RNA
ms	millisecond
MW	Molecular Weight
N	Nonality
NaCl	Sodium Chloride
NaOAc	Sodium Acetate
NaOH	Natrium Hydroxide
NEC	Non-Embryogenic Callus
ng	nanogram

nm	nanometer
nM	nanomolar
nt	nucleotide
N-terminal	Amino terminal
OD	Optical density
ORF	Open Reading Frame
PAGE	Polyacrylamide Agarose Gel Electrophoresis
PCR	Polymerase Chain Reaction
PE	Proembryo
pfu	plaque forming unit
Poly A^+RNA	Polyadenylated RNA
PPO	2,5-Diphenyloxazole
RE	Restriction enzyme
RNA	Ribonucleic Acid
RNase	ribonuclease
rpm	revolution per minute
rRNA	ribosomal RNA
RT	Room Temperature
SDS	Sodium Dodecyl Sulphate
sec	seconds
STE	Sodium-Tris-EDTA
T	Thiamine
TAE	Tris-Acetate-EDTA
TBE	Tris-Borate-EDTA
TCA	Trichloroacetic Acid

TE	Tris-EDTA
TEMED	N, N, N', N'-Tetramethylethylenediamine
Tm	Annealing temperature
U.P.M	University Putra Malaysia
UV	ultraviolet
v/v	volume per volume
w/v	weight per volume
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranose
XL1-B	XL1-Blue MRF'

CHAPTER I

INTRODUCTION

Oil palm (*Elaeis guineensis* Jacq.) was first introduced into Malaysia from Central and Western Africa in the early 1870's (Hartley, 1977). It is currently the second major source of vegetable oil in the world after soybean (Mielke, 1996). The oil palm belongs to the family Palmaceae and the genus *Elaeis*. There are two important species in the genus *Elaeis*, they are *E. guineensis* and *E. oleifera*. The former is of African origin and the latter is of South American origin. This diploid monocotyledon is persistent in the dominance of the single vegetative apex, hence producing no adventitious or auxillary shoots.

Prior to the advent of tissue culture, there was no known reliable method of vegetative propagation of oil palm (Wong *et al*., 1996). Therefore tissue culture is an important aspect in the development of the oil palm industry (Cheah & Wooi, 1993), especially in the generation of superior and uniform oil palm planting materials. Pioneering work in oil palm tissue culture was carried out by Staritsky (1970) and Rabechault *et al*., (1970), and the first successful cloning of mature palms was reported by Rabechault *et al*., (1972) and Corley *et al*., (1977). However, tissue culture has its limitation. The current success rate of callogenesis of oil palm explants remains low, at about 20 percent, while the average rate of embryogenesis is only 6 percent as reported by most tissue culture laboratories (Wooi, 1995). Despite the economic importance of oil palm tissue culture, little is known about the chemical characteristics and molecular changes associated with callogenesis and embryogenesis.

In plants, a number of approaches have been used to understand the complexity of gene expression and interaction. The generation of large numbers of partial cDNA sequences, or expressed sequence tags (ESTs), has provided a method to sample a large number of genes from an organism. Since its introduction in 1991, the generation of expressed sequence tags (ESTs) by single-run partial sequencing has proven to be a rapid and efficient way of obtaining information on mRNA expression patterns from a wide variety of tissues, cell types, or developmental stages (Adam *et al.*, 1995; Kawamoto *et al.*, 1996). It was found that as little as 150 to 400 bases is sufficient for a match to sequences resident in the databases (Adams *et al.*, 1991). The results of these comparisons are used as a guide to assign putative identification to the ESTs.

With respect to tissue culture, it has been suggested that many of the functional genes necessary for embryogenesis are expressed at an early stage from an unorganized growth state (Sung & Okimoto, 1983; Giuliano & Terzi, 1985; Wilde *et al.*, 1988; Zimmerman, 1993). Thus a comparison of gene expression patterns in embryogenic callus and non-embryogenic callus using statistical analysis of ESTs would give us an insight of not only the embryogenic potency but also the regeneration mechanism of the plant. Therefore this study is aimed at understanding the gene expression pattern in non-embryogenic callus of oil palm. This is carried out by the generation and statistical analysis of ESTs from non-embryogenic callus of oil palm. This study hopes to generate some insights into the complexity of genes expressed in non-embryogenic callus of oil palm.

CHAPTER II

LITERATURE REVIEW

2.1 The Oil Palm

2.1.1 Origin and Distribution

It is thought that the centre of origin of the oil palm is the tropical rain forest region of West and Central Africa. The plant later spread from Senegal to Angola in a wide belt along the coast and interior of the Congo River (Hartley, 1977). The oil palm grows and reproduces readily in a wild and semi-wild state and its successful distribution was due to the influence of human activities (Ascenco, 1966). The sporadic occurrence on the East African coast was most likely due to Arab slave traders.

The palm fruits were taken by the early slave traders in the 16th Century to the New World, where it became established first in Bahia, Brazil. Favorable climate and shifting of cultivation practices in Brazil resulted in extensive areas of semi-wild palm groves. Oil palm seedlings including *Elaeis guineensis* were grown in European conservatories in the 18th Century, and, in the following century, it was brought to Calcutta as ornamental plants. The original introduction to the East was from the island of Mauritius. Four seedling trees were then established in 1848 in the Botanic Garden at Bogor in Java, which were until the 1950s the original parent to all the oil palms in Indonesia and Malaya (Whitmore, 1979).

Oil palm was first introduced into Malaya in 1875 (Iman, 1984) and the first commercial plantings for production of edible oil were in 1917. By 1938, there were 30, 000 ha of land in Malaya already planted with oil palm. After World War II, production of palm oil increased considerable due to favourable economics. Prices of both oil and kernels reached four or five times their pre-war level and this remained comparatively steady for the next 20 years (Hartley, 1977). Due to the higher profitability and demands caused by the oil's wide use in industrial applications, rubber plantations were slowly but steadily being replaced by oil palm plantation in addition to new plantation areas developed (Moll, 1987).

2.1.2 Botany

2.1.2.1 Classification

The genus *Elaeis* belongs to the subfamily Cocoideae and to the palm family, Palmae or Arecaceae, which is an important member of the monocotyledonous group. It has a basic chromosome number of n=16 and classified in the order Spadiflorae. The word, *Elaeis* is derived from the Greek word 'elaion', oil, while the specific name *guineensis* shows the palm origin, which is the Guinea coast. The fruits of the oil palm vary widely and the classifications are based on variations of the internal structure of the fruits. The examples of these classifications are *Dura* (shell 2 to 8 mm thick), *Pisifera* (shell-less), and *Tenera* (Shell 0.5 to 4 mm thick). The tenera fruit form is a hybrid of the dura and pisifera forms, often designated as D x P. Since the late 1950s most commercially planted palms in Malaysia are Tenera (Whitmore, 1979). The South American oil palm, *Elaeis oleifera* (syn. *Elaeis melanococca*) is another species in the genus *Elaeis*. However due to its low oil yield, it is of little economic importance except